



Exploration visuelle et interactive d'une large collection d'images en libre accès

Frédéric Rayar, Sabine Barrat, Fatma Bouali, Gilles Venturini

► To cite this version:

Frédéric Rayar, Sabine Barrat, Fatma Bouali, Gilles Venturini. Exploration visuelle et interactive d'une large collection d'images en libre accès. 2015. hal-01168720

HAL Id: hal-01168720

<https://hal.science/hal-01168720>

Submitted on 26 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploration visuelle et interactive d'une large collection d'images en libre accès

Frédéric Rayar*, Sabine Barrat*
Fatma Bouali*,** Gilles Venturini *

*Université François-Rabelais de Tours, Laboratoire d'Informatique
64 avenue Jean Portalis, 37200 Tours, France,
frederic.rayar@univ-tours.fr, sabine.barrat@univ-tours.fr, gilles.venturini@univ-tours.fr

**Université de Lille2, IUT, Dpt STID
25-27 Rue du Maréchal Foch, 59100 Roubaix, France,
fatma.bouali@univ-lille2.fr

Résumé. Cet article décrit l'approche et les résultats de travaux préliminaires sur l'exploration visuelle et interactive d'une large collection d'images en libre accès (Open Access images). Cette exploration s'appuie sur des descripteurs visuels et structure les images dans un graphe de proximité de manière incrémentale, afin de pouvoir traiter de grandes collections d'images. L'approche utilisée consiste à : (i) collecter des images et des métadonnées, (ii) extraire des descripteurs, (iii) structurer des données dans le graphe et (iv) visualiser le graphe obtenu à l'aide d'algorithmes existants de dessin de graphes.

1 Introduction

La gestion et l'analyse de grandes quantités de données sont des défis majeurs actuels de la communauté scientifique. On retrouve communément cette problématique à travers l'expression « *big data* » ou encore « *data deluge* ». Ce défi se décline en plusieurs aspects tels que la collecte et le stockage efficace des données, mais aussi leur analyse, à savoir rechercher, résumer, visualiser et explorer ces données. Parmi cette immense quantité de données accessibles, on distingue les données ouvertes (*open data*), produites par des instituts publics ou privés. Ce domaine d'application commence progressivement à attirer l'attention des chercheurs, tant par l'aspect quantitatif que l'aspect non structuré des données étudiées.

Dans les travaux préliminaires présentés dans cet article, nous nous intéressons particulièrement aux images *libres de droits* (*open access images*), qui, à notre connaissance, n'ont pas fait l'objet de travaux de recherche. En effet, de nombreux instituts tels que les galeries d'art, les bibliothèques, les archives ou encore les musées, continuent leurs efforts pour s'adapter à l'ère numérique, et mettent en place des projets de numérisation et de mise en ligne d'œuvres historiquement inestimables et parfois peu connues ou oubliées. On peut notamment citer, parmi les instituts les plus connus, le Metropolitan Museum of Art (MET) de New York¹, la

1. <http://www.metmuseum.org/>

British Library² de Londres, ou encore le Rijksmuseum d'Amsterdam³. Outre l'aspect quantitatif des images à traiter (de quelques milliers à des millions), nous portons ici notre attention sur une exploration d'une large collection d'images s'appuyant sur des descripteurs visuels, afin de permettre une navigation par le contenu.

La suite de l'article est présentée de la manière suivante : la section 2 présente un état de l'art de la visualisation de grandes collections d'images. L'approche utilisée dans le cadre de ces travaux préliminaires est ensuite détaillée dans la section 3. Puis, les résultats des expériences conduites sont présentés et discutés dans la section 4. Enfin, la section 5 résume les contributions apportées et répertorie les perspectives de ces travaux.

2 Etat de l'art

La dernière décennie a vu la réduction du coût des appareils photos, des webcams et des scanners, mais aussi celui des supports de stockage. De fait, la quantité d'images capturées par tout un chacun a explosé, qu'elles soient générées dans le cadre privé, commercial ou dans celui de projets de numérisation. De plus, l'avènement d'Internet a accentué le fait que le nombre d'images mises en ligne croît de manière exponentielle, notamment avec les sites d'instituts, de e-commerces ou encore les réseaux sociaux.

Ainsi, les collections d'images devenant conséquentes, il est apparu nécessaire de proposer des paradigmes pour les visualiser. Une analyse plus complète de cette problématique, des difficultés sous-jacentes et des solutions existantes peut être trouvée dans l'article de Plant et Schaefer (2011). On peut classer les solutions proposées de la manière suivante :

- *les visualisations temporelles et géographiques* : elles se basent sur les informations temporelles (*e.g.* timestamp) ou géographiques (*e.g.* geotags) et s'appuient sur le fait que les utilisateurs sont accoutumés à interagir avec des frises chronologiques et des planisphères (Nöllenburg (2007), Jaffe et al. (2006)) ;
- *les visualisations basées sur des mots-clés* : ces derniers peuvent être récupérés directement ou extraits à partir des descriptions et annotations des images ; une visualisation s'appuyant sur une liste ou un nuage de mots permet ensuite d'explorer la collection d'images ;
- *les visualisations basées sur les similarités entre images* ; ici les images de la collection sont organisées selon ces similarités ; on détaille plus bas ce type de visualisation.

Il est à noter que les deux premières catégories de visualisation ne sont exploitables que si l'on dispose, pour chaque image, des informations temporelles, géographiques ou des mots-clés, ce qui n'est pas forcément le cas dans certaines collections d'images. La dernière catégorie, quant à elle, ne s'appuie que sur les informations inhérentes aux images, *i.e.* les pixels ou autres descripteurs visuels. L'idée est donc de décrire une image à l'aide d'un ensemble de d valeurs : elle peut alors être considérée comme un point dans l'espace de descripteurs de dimension d . L'ensemble des images de la collection est ainsi positionné dans cet espace et on cherche

2. <http://www.bl.uk/>

3. <https://www.rijksmuseum.nl/en>

dès lors à représenter la structure du *nuage de points* correspondant. Quatre approches ont été proposées :

- (i) *l'approche projection* : on cherche ici à projeter l'espace originel de dimension d dans un espace de dimension 2 (éventuellement 3), de manière à permettre à l'esprit humain de percevoir la structure des points. La projection se doit de préserver la topologie des points observée dans l'espace de dimension d . Parmi les méthodes les plus connues, on peut citer l'*Analyse en Composantes Principales* (PCA) ou encore le positionnement multidimensionnel (MDS), utilisé dans (Rubner et al. (1997)).
- (ii) *l'approche graphe* : un graphe est construit, dans lequel les nœuds correspondent aux images, et les arcs entre deux nœuds représentent la similarité entre les deux images associées. Le problème se résume donc à déterminer quels arcs créer et quels algorithmes de dessins de graphes utiliser pour visualiser le graphe obtenu. Différentes solutions ont été envisagées, s'appuyant principalement sur des graphes kNN (Heesch et Rüger (2004)).
- (iii) *l'approche clustering* : l'idée clé est de réduire le nombre d'images à afficher dans un premier temps. Les images similaires sont donc regroupées ensemble afin de former des groupes homogènes et ce de manière hiérarchisée. Une visualisation multi-niveau peut ainsi être exploitée comme dans (Krishnamachari et Abdel-Mottaleb (1999)). Il est à noter que cette approche de clustering peut aussi être utilisée sur des informations temporelles, géographiques ou des mots-clés.
- (iv) *l'approche vues multiples et coordonnées* : on se base ici sur des visualisations classiques de l'analyse de données telles que les histogrammes, les scatterplots ou encore les coordonnées parallèles. La valeur ajoutée réside dans le fait de proposer une visualisation multiple, simultanée et synchronisée de ces différentes techniques visuelles (Matković et al. (2009)).

Suite à nos précédents travaux sur les *graphes de proximité*, nous nous appuyerons dans le cadre de ces travaux préliminaires sur l'approche graphe, afin de visualiser une large collection d'images en libre accès.

3 Approche proposée

Notre approche peut être résumée comme suit (*cf.* Figure 1) :

1. collecte des images depuis le portail en ligne d'un institut (les métadonnées associées sont aussi récupérées en vue de travaux futurs) ;
2. extraction de descripteurs visuels : les propositions de la norme MPEG-7 sont exploitées ici ;
3. structuration des données à l'aide d'un graphe de proximité : le graphe des voisins relatifs (RNG) ;
4. visualisation du graphe obtenu à l'aide des logiciels Tulip⁴ et Gephi (Bastian et al. (2009)).

4. <http://tulip.labri.fr>

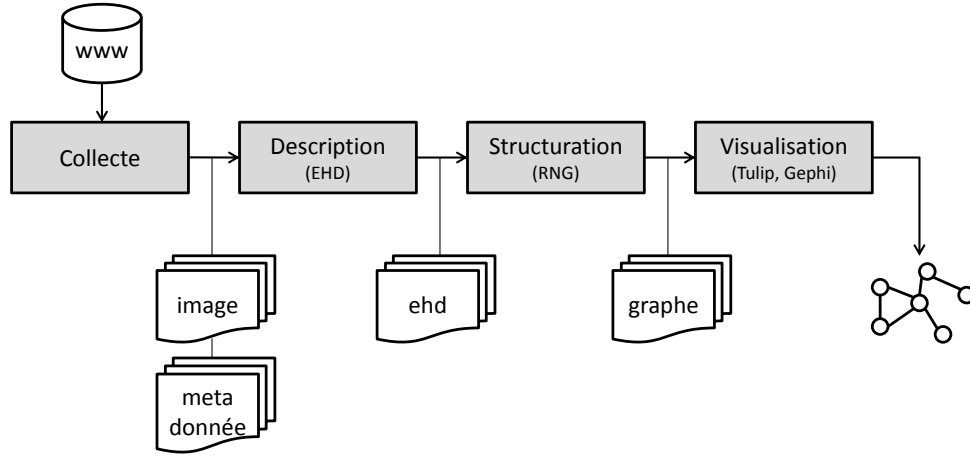


FIG. 1: Workflow de l'approche proposée.

3.1 Collecte des images

Ce travail étant prospectif, nous proposons de récupérer l'ensemble de la collection d'images à étudier en local, avec les métadonnées associées. Bien qu'il faille noter les efforts de certains instituts (*e.g.* la British Library) pour offrir un accès simple à la totalité de leur collection, au travers d'API (*e.g.* l'API Flickr), la majeure partie de ces collections n'est pas directement accessible. Aussi, il est nécessaire de créer un outil personnalisé, de type aspirateur de sites web, pour récupérer l'ensemble de ces données.

3.2 Extraction de descripteurs visuels

MPEG-7 (Pereira et Koenen (2001)) est une norme de description de contenu multimédia qui a pour objectif de faciliter les opérations d'indexation et recherche sur des documents multimédias. On retrouve dans la définition de cette norme un ensemble de descripteurs visuels permettant la description d'images ou de vidéos, à partir des informations de couleur, texture et forme. Un grand nombre d'images de la collection étudiée étant en niveaux de gris, nous avons décidé, dans un premier temps, de ne pas utiliser de descripteurs de couleur. Ainsi, dans ce travail préliminaire, nous nous limitons à un descripteur de texture, nommé les *edge histogram descriptors* (EHD)(MPEG (2000)).

Ce descripteur représente la distribution de cinq types de contours dans des régions de l'image. La Figure 2a présente les types de contours considérés. Ainsi, une image est divisée en 4x4 cellules disjointes (Figure 2b) et, pour chaque cellule, un histogramme de la distribution de chaque type de contours est créé. On a donc $4 \times 4 \times 5 = 80$ valeurs permettant de décrire chacune des images. Il est à noter que certains travaux ont été proposés pour améliorer cette définition basique des EHD pour la comparaison d'images (Won et al. (2002)).

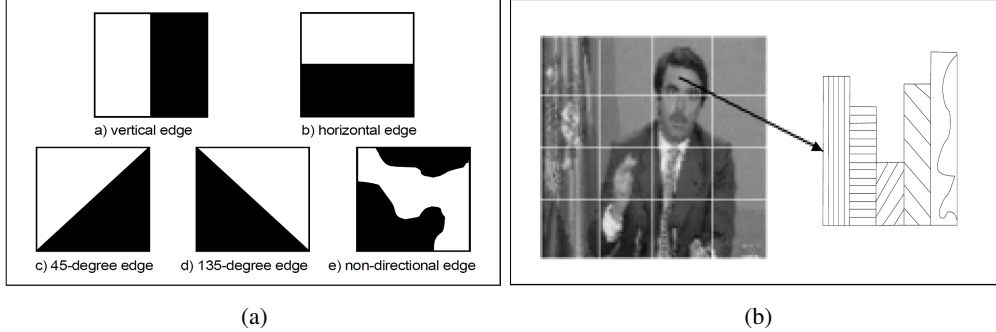


FIG. 2: Descripteurs à base d'histogramme de contour : (a) les 5 types de contours considérés et (b) le calcul d'un histogramme de contour pour chaque cellule. Illustrations tirées de Won et al. (2002).

3.3 Structuration des données

Afin de structurer nos données, nous avons choisi le *graphe des voisins relatifs* (Toussaint (1980)). En effet, le graphe est une structure simple qui permet de représenter différents types de données : les nœuds correspondent aux données et les arcs aux relations entre ces données. Les propriétés de cette structure ont été abondamment étudiées, et continuent de l'être, dans le domaine de la théorie des graphes. En particulier, les *graphes de proximité* constituent une famille de graphes permettant d'extraire une structure inhérente aux données étudiées : deux nœuds sont reliés par un arc s'ils sont suffisamment proches pour être considérés comme des voisins. On note parmi les graphes de proximité les plus connus, l'arbre recouvrant de poids minimal (MST) (Boruvka (1926)), le graphe kNN (Preparata et Shamos (1985)), le RNG (Toussaint (1980)), le graphe de Gabriel (Gabriel et Sokal (1969)), ou encore le graphe de Delaunay (Delaunay (1934)). Le RNG ayant la particularité d'être le plus petit graphe de proximité connexe (*i.e.* il existe un chemin entre tout couple de nœud), c'est celui-ci qui sera utilisé dans nos travaux. En effet, ce type de structure permettra une exploration de la collection par le contenu, en se déplaçant de proches en proches. Par définition, deux nœuds p et q sont des voisins relatifs s'il n'existe aucun autre nœud r plus proche de p et de q qu'ils ne le sont entre eux. Ainsi, on a :

$$E(RNG) = \{\overline{pq}, p, q \in D, p \neq q, \delta(p, q) \leq \max(\delta(p, r), \delta(q, r)), \forall r \in D \setminus \{p, q\}\},$$

avec D , l'ensemble des nœuds considérés, $E(RNG)$, l'ensemble des arcs du RNG et $\delta : D \times D \rightarrow \mathbb{R}$, une distance. La Figure 3 illustre la notion de voisinage relatif.

L'inconvénient majeur du RNG est la complexité de sa construction en $O(n^3)$. De fait, une construction classique semble difficilement applicable à de grands jeux de données. Afin de pallier cette limite, nous avons étudié et amélioré un paradigme de construction *incrémentale* du RNG, initialement proposé par Hacid et Zighed (2005). Ce travail fait l'objet d'une soumission à une conférence internationale. Ce paradigme propose, lors de l'insertion d'un nouveau nœud dans un RNG existant, de venir mettre à jour uniquement une région pertinente du graphe, permettant ainsi de ne pas reconsidérer les parties du graphe « éloignées » du nouveau nœud inséré.

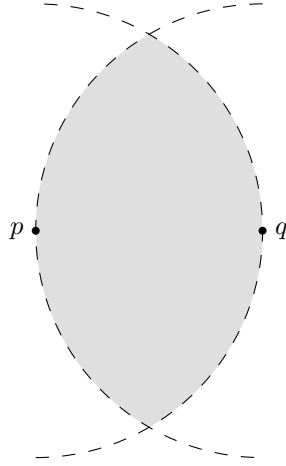


FIG. 3: Voisinage relatif de deux points $p, q \in \mathbb{R}^2$ (région grisée). Si cette région ne contient aucun autre point, alors p et q sont des voisins relatifs.

3.4 Visualisation

Afin d'observer la cartographie des données, et permettre de révéler d'éventuels groupes homogènes (*clusters*) ou des données aberrantes (*outliers*) par un utilisateur, nous utilisons une représentation visuelle du graphe. Les logiciels Tulip (version 4.6) et Gephi (version 0.8.2 beta) permettent de telles visualisations. Ces outils proposent un ensemble de méthodes de dessin de graphes générant différentes représentations visuelles, mais aussi un ensemble de fonctionnalités et d'interactions standards permettant d'explorer les graphes.

4 Expériences

4.1 Données

La collection d'images de la *National Gallery of Art* (NGA) de Washington⁵ a été utilisée dans le cadre de ces travaux préliminaires. Elle est constituée de plus de 43000 images en libre accès, notamment de tableaux, de photographies de sculptures ou encore de dessins. La Figure 4 présente quelques images de cette collection.

Ces images n'étant pas directement accessibles, nous avons dû créer un aspirateur d'images dédié. Dans un premier temps, l'ensemble des « asset id » a été récupéré. Selon la définition du NGA, un « asset id » correspond à l'identifiant unique d'une image. Des requêtes HTTP ont ensuite été exécutées pour télécharger les images en taille originale. Les métadonnées associées ont quant à elles été extraites du code source HTML, puis converties au format XML. 43721 images et leurs métadonnées ont ainsi été récupérées. Pour chacune des images, le descripteur EHD a été calculé et stocké dans un fichier afin de ne pas avoir à accéder de

5. <https://images.nga.gov>



FIG. 4: Exemples d'images de la collection de la NGA : (a) un tableau, (b) une photographie de sculpture et (c) un dessin.

nouveau aux images durant l'étape de construction du graphe. Une matrice 43721×81 a ainsi été obtenue, décrivant la collection d'images du NGA (80 valeurs pour le descripteur, plus l'id de chaque image). Notre algorithme de construction incrémentale du RNG a traité ces données et généré un graphe composé de 117925 arcs (90 minutes sur un Intel® Xeon® CPU W3520 à 2.67Ghz). Enfin, plusieurs stratégies de dessin de graphes ont été utilisées afin d'observer différentes représentations visuelles : l'algorithme OpenOrd (Martin et al. (2011)) et l'algorithme multi-niveau de Yifan Hu's (Hu (2005)) sur Gephi, ainsi que l'algorithme FM³ sur Tulip (Hachul et Jünger (2005)). Les Figures 5a, 5b et 6a présentent respectivement les tracés de graphes obtenus avec ces algorithmes.

4.2 Résultats

Forts de ces quelques visualisations, nous pouvons émettre un certain nombre d'observations et d'hypothèses. Premièrement, on distingue dans la première visualisation (Figure 5a) obtenue avec Gephi et l'algorithme OpenOrd, un groupe homogène très dense que l'on pourrait qualifier de « noyau », autour duquel un certain nombre de nœuds gravitent de manière désordonnée. Dans la seconde visualisation (Figure 5b) obtenue avec Gephi en exploitant l'algorithme multi-niveau de Yifan Hu, on retrouve le noyau précédemment observé, ainsi que des nœuds présents en périphérie, qui, semblent cette fois s'organiser en petits groupes homogènes.

La visualisation obtenue avec le logiciel Tulip et l'algorithme FM³ (Figure 6a) présente elle aussi un tracé de type noyau autour duquel gravitent des nœuds. Cependant, dans cette visualisation, le noyau semble contenir très peu de nœuds, et les nœuds de la périphérie s'organisent tels des filaments autour de ce noyau. Ce résultat ne semble donc pas en adéquation avec les deux précédentes observations. Néanmoins, les outils d'interaction fournis par Tulip, permettent de manipuler le graphe et de découvrir que cette vue globale du graphe est légèrement

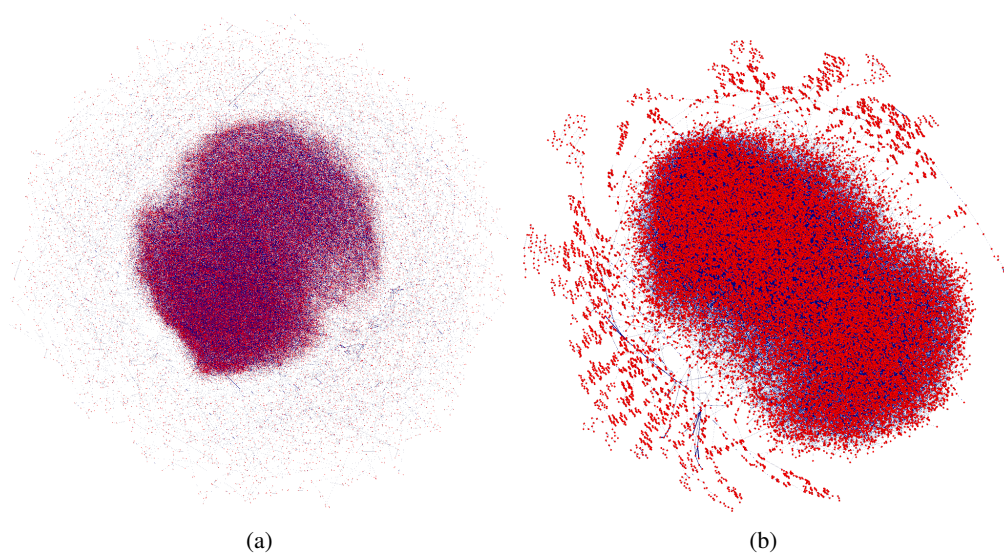


FIG. 5: Visualisation du RNG obtenu avec Gephi en utilisant (a) l'algorithme OpenOrd et (b) l'algorithme multi-niveau de Yifan Hu.

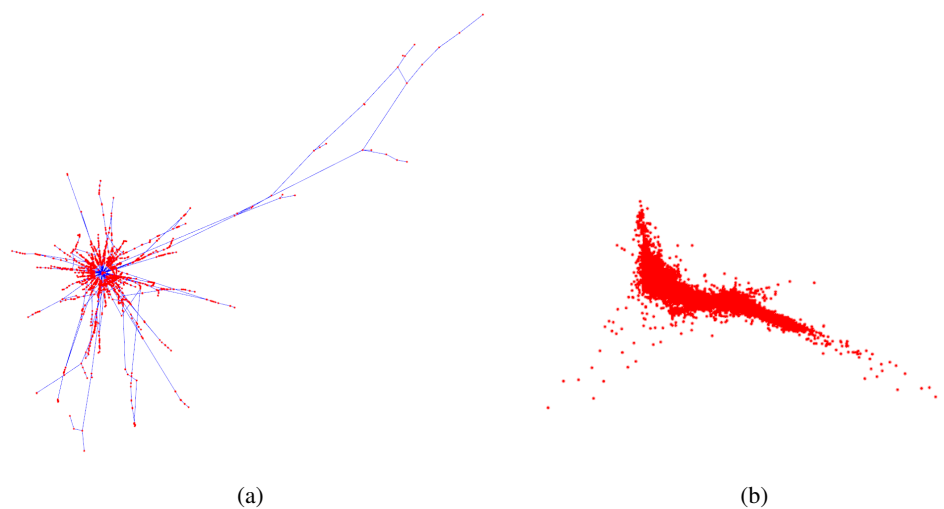


FIG. 6: Visualisation du RNG obtenu avec Tulip et l'algorithme FM^3 (a) dans sa totalité et (b) en zoomant au niveau du noyau.

trompeuse. En effet, le noyau observé contient en réalité 91% des nœuds : on retrouve ainsi la structure de noyau homogène très dense, avec une organisation plus structurée des nœuds en périphérie. La Figure 6b présente un zoom sur le noyau dans le tracé de FM³ dans Tulip.

Ainsi, ces différentes visualisations du graphe et les observations effectuées permettent deux axes d'exploration :

- le fait qu'une grande majorité de nos images (91%) semblent être proches, au sens de la distribution des contours, peut pousser à remettre en question le choix du descripteur de texture utilisé, peu discriminant pour la masse ;
- néanmoins, au vu des tracés du graphe, nous pouvons aussi orienter l'exploration de la collection d'images vers une investigation plus avancée sur ces petits groupes homogènes qui gravitent autour du noyau, et qui pourraient révéler une information intéressante.

5 Conclusion et travaux futurs

Nous avons présenté ici l'un des premiers travaux, à notre connaissance, sur l'exploration visuelle d'une large collection d'images en libre accès, ces dernières pouvant amener un certain nombre de spécificités et difficultés.

Un outil de type aspirateur de sites a été développé afin de récupérer les images et les métadonnées associées. Un descripteur de type texture (EHD) a ensuite été calculé pour chacune des images. Cette description des images a été utilisée par notre algorithme de construction incrémentale du RNG qui permet de traiter de grandes collections d'images. Enfin, la visualisation du graphe obtenu à l'aide de logiciels existants a permis d'explorer de manière interactive la collection d'images.

L'approche utilisée dans ce travail préliminaire et les premiers résultats obtenus, bien qu'intéressants, nous permettent de proposer différentes pistes d'améliorations telles que :

- une expérimentation d'autres descripteurs visuels tels que des descripteurs de couleur et forme, une combinaison de ces descripteurs ou encore la combinaison de descripteurs visuels avec des métadonnées,
- l'étude de collections plus grandes (*e.g.* la British Library avec plus d'un million d'images mises en ligne) : cela semble néanmoins nécessiter le développement d'une visualisation multi-niveau dédiée, pour permettre une meilleure exploration du graphe et des interactions plus fluides,
- une étape de clustering sur le graphe qui permettrait une meilleure analyse et visualisation multi-niveau.

Références

- Bastian, M., S. Heymann, et M. Jacomy (2009). Gephi : An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.
- Boruvka, O. (1926). O Jistém Problému Minimálním (About a Certain Minimal Problem) (in Czech, German summary). *Práce Mor. Přírodoved. Spol. v Brně III* 3.

- Delaunay, B. N. (1934). Sur la sphère vide. *Bulletin of Academy of Sciences of the USSR* 7, 793–800.
- Gabriel, R. K. et R. R. Sokal (1969). A New Statistical Approach to Geographic Variation Analysis. *Systematic Zoology* 18(3), 259–278.
- Hachul, S. et M. Jünger (2005). Drawing large graphs with a potential-field-based multilevel algorithm. In J. Pach (Ed.), *Graph Drawing*, Volume 3383 of *Lecture Notes in Computer Science*, pp. 285–295. Springer Berlin Heidelberg.
- Hacid, H. et D. A. Zighed (2005). An effective method for locally neighborhood graphs updating. In *DEXA*, pp. 930–939.
- Heesch, D. et S. Rüger (2004). Nn^k networks for content-based image retrieval. In *Advances in Information Retrieval*, Volume 2997, pp. 253–266. Published in : S. McDonald and J. Tait (Eds.) : ECIR 2004, LNCS 2997, pp. 253–266, 2004.
- Hu, Y. F. (2005). Efficient and high quality force-directed graph drawing. *The Mathematica Journal* 10, 37–71.
- Jaffe, A., M. Naaman, T. Tassa, et M. Davis (2006). Generating summaries and visualization for large collections of geo-referenced photographs. In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, MIR '06, pp. 89–98. ACM.
- Krishnamachari, S. et M. Abdel-Mottaleb (1999). Image browsing using hierarchical clustering. In *Computers and Communications, 1999. Proceedings. IEEE International Symposium on*, pp. 301–307.
- Martin, S., W. M. Brown, R. Klavans, et K. W. Boyack (2011). Openord : an open-source toolbox for large graph layout. *Proc. SPIE* 7868, 786806–786806–11.
- Matković, K., D. Gračanin, W. Freiler, J. Banova, et H. Hauser (2009). Large image collections - comprehension and familiarization by interactive visual analysis. In *Smart Graphics*, Volume 5531 of *Lecture Notes in Computer Science*, pp. 15–26. Springer Berlin Heidelberg.
- MPEG (2000). Iso/ieci/jtc1/sc29/wg11 - core experiment results for edge histogram descriptor (ct4)), mpeg document m6174. Technical report.
- Nöllenburg, M. (2007). Geographic visualization. In *Human-Centered Visualization Environments*, Volume 4417 of *Lecture Notes in Computer Science*, pp. 257–294. Springer Berlin Heidelberg.
- Pereira, F. et R. Koenen (2001). MPEG-7 : A standard for multimedia content description. *Int. J. Image Graphics* 1(3), 527–546.
- Plant, W. et G. Schaefer (2011). Visualisation and browsing of image databases. In *Multimedia Analysis, Processing and Communications*, Volume 346 of *Studies in Computational Intelligence*, pp. 3–57. Springer Berlin Heidelberg.
- Preparata, F. P. et M. I. Shamos (1985). *Computational Geometry : An Introduction*. New York, NY, USA : Springer-Verlag New York, Inc.
- Rubner, Y., L. Guibas, et C. Tomasi (1997). The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop*, pp. 661–668.

- Toussaint, G. T. (1980). The relative neighbourhood graph of a finite planar set. *Pattern Recognition* 12, 261–268.
- Won, C. S., D. K. Park, et Y. S. Jeon (2002). Efficient use of mpeg-7 edge histogram descriptor. *ETRI Journal* 24(1), 23–30.

Summary

This paper describes the principles and present the first results of a preliminary study on the visual and interactive exploration of large Open Access images collections. The mentioned exploration relies on visual descriptors and structures the images incrementally in order to handle large collections. The principles can be summarized in the following steps: *(i)* images and metadata gathering, *(ii)* descriptors extraction, *(iii)* data structuring in a proximity graph and *(iv)* graph visualisation using existing graph drawing methods.